DOCUMENT RESUME

ED 198 151 TM 810 073

AUTHOR de Gruijter, Dato N. M.
TITLE Empirical Bayes Estimation in the Rasch Model: A Simulation.
PUB DATE Jun 80
NOTE 13p.: Paper presented at the European Meeting of the Psychometric Society (Groningen, Netherlands, June 19-21, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bayesian Statistics: Comparative Analysis: Goodness of Fit: Item Sampling: *Latent Trait Theory: *Mathematical Models: *Maximum Likelihood Statistics: Population Distribution
IDENTIFIERS *Computer Simulation: Normality Tests: *Rasch Model

ABSTRACT

In a situation where the population distribution of latent trait scores can be estimated, the ordinary maximum likelihood estimator of latent trait scores may be improved upon by taking the estimated population distribution into account. In this paper empirical Bayes estimators are compared with the liklihood estimator for three samples of 300 cases within the context of the Rasch model. The empirical Bayes estimates varied more than the likelihood estimates, due to the fact that not only item parameters, but also population distribution parameters had to be estimated from the sample data, the largest difference being 0.09 for a score equal to 20. The results were based on a computer simulation for which the model and distributional assumptions were known to be correct. For real data, there is a question of appropriateness of the distributional assumptions--apart from the question of fit of the Rasch Model. The normality assumption can be tested by means of the test of fit proposed by Andersen and Madsen. If the normality assumption turns out to be inadequate, other distributions may be fitted to the data using an approach similar to that proposed in this paper. (Author/RL)

Educational Research Center
University of Leyden

Empirical Bayes estimation in the Rasch model: a simulation *

Dato N.M. de Gruijter
Educational Research Center
University of Leyden

In a situation where the population distribution of latent trait scores can be estimated, one may improve upon the ordinary maximum likelihood estimator of latent trait scores by taking the estimated population distribution into account. In the present paper empirical Bayes estimators are compared with the likelihood estimator within the context of the Rasch model. The data are simulated, with latent trait scores generated from a normal distribution.

Introduction

In a latent trait model person parameters or abilities can be estimated by the method of maximum likelihood, given the item parameters or estimates of the item parameters. The ML-method has some disadvantages, however. First, person parameters cannot be estimated for persons with a zero or perfect score, the estimates tending to minus and plus infinity, respectively; in the three parameter logistic model also other score patterns have no unique maximum of the likelihood equation (Samejima, 1973). Secondly, in case the persons can be regarded as randomly sampled from some population of persons, an empirical Bayes estimate of persons' abilities can be obtained which has a smaller expected mean squared error and is preferable for that reason; in the same way interval estimates for abilities can be obtained which compare favorably with intervals obtained through the information function.

* Paper presented at the European Meeting of the Psychometric Society, Groningen, June 19 thru 21, 1980. The author is indebted to Eric A. Bakker who wrote the computer programs.

Bayesian and empirical Bayes procedures have been proposed before within the context of latent trait theory. Birnbaum (1969) and - in connection with tailored testing - Owen (1969; see also Jensema, 1974) have proposed a Bayesian analysis for the logistic and normal ogive latent trait models. Leonard (1972) has presented a Bayesian analysis for the binomial model using the 'log-odds' transformation; his model can be regarded as a special case of the Rasch model with all item parameters equal to zero. An empirical Bayes point estimate of ability has been derived by Meredith and Kearns (1973) for the multiplicative version of the Rasch model; the advantage of their approach is that no specific assumptions with respect to the population distribution have to be made and that the estimation of the population distribution can be bypassed. Sanathanan and Blumenthal (1978) have presented an empirical Bayes procedure for the additive Rasch model; they assume that the person parameter density belongs to a specific family of densities. In the present paper an empirical Bayes procedure for the additive version of the Rasch model is proposed, assuming normally distribution abilities.

The posterior distribution in the Rasch model

The additive Rasch model reads

(1) $P(u_i=1|\theta)=\exp(\theta-b_i)/\left[1+\exp(\theta-b_i)\right]$

where θ is the person parameter or ability, b_i is the item parameter or difficulty of item i and $P(u_i=1|\theta)$ is the probability of a correct answer on item i, given ability θ. Given the item parameters $\underset{\sim}{b}=(b_1,b_2,\ldots,b_n)$ or good item parameter estimates for the items of a n-item test, the likelihood of having t items correct equals

(2) $L_{\underset{\sim}{b}}(t|\theta)=\gamma_t \exp(t\theta) \left\{\prod_{i=1}^{n} \left[1+\exp(\theta-b_i)\right]\right\}^{-1}$

where γ_t is the elementary symmetric function of order t of the n factors $\exp(-b_i)$.

Let us assume that the person parameters can be considered as randomly sampled from a population distribution $g(\theta)$, i.e., $g(\theta)$ is the prior distribution for the θ's. Using Bayes' rule we obtain the posterior distribution of θ

$$(3)\quad f(\theta|t) = \frac{f(t|\theta)g(\theta)}{f(t)}$$

$$= \frac{1}{f(t)} \cdot \gamma_t \exp(t\theta) \left\{ \prod_{i=1}^{n} \left[1+\exp(\theta-b_i) \right] \right\}^{-1} g(\theta)$$

where

$$(4)\quad f(t) = \int \gamma_t \exp(t\theta) \left\{ \prod_{i=1}^{n} \left[1+\exp(\theta-b_i) \right] \right\}^{-1} g(\theta)d\theta \, .$$

Equation (3) is similar to Birnbaum's (1969) Equation (4). In the present paper it is assumed that $g(\theta)$ is a normal distribution $\phi(\theta|\mu,\sigma^2)$.

The above model is not quite realistic in that a known population distribution is assumed. The more general approach is to assume that prior information on μ and σ^2 is available and to combine this information with data from the sample of persons in order to obtain posterior estimates of μ and σ^2. In an empirical Bayes approach only sample information is used in order to obtain the parameters of the population distribution.

The normal population distribution

In the logistic and normal ogive models the ability scale has interval characteristics, i.e. the models are invariant up to a linear transformation. For this reason one can choose the representation for which θ is $N(0,1)$. Item parameters have to be estimated on this particular scale. Bock and Lieberman (1970) present an estimation procedure for the two-parameter normal ogive model, Bock (1972) presents one for the logistic model without guessing parameter. An heuristic procedure for the three-parameter ogive model has been proposed by Urry (1976).

The same could be done in connection with the Rasch model if one allows the common discrimination parameter to be unequal to one. We will, however, confine ourselves to the Rasch model proper in which only translations of the ability scale are allowed.

For the Rasch model a procedure for the estimation of μ and σ^2 of the population distribution has been presented by Andersen and Madsen (1977). They start with the marginal frequencies $f(t)$, $t=0,1,\ldots,n$, from (4) assuming accurately estimated item parameters. The likelihood of N_t scores t for $t = 0,1,\ldots,n$ in a sample of $N = \sum_{t=0}^{n} N_t$ persons from the population equals

$$\text{(5)}\quad L_{\underset{\sim}{b},\phi} = \sum_{t=0}^{n} \left[f(t|\mu,\sigma^2)\right]^{N_t}$$

where ϕ denotes the normal frequency distribution.

Estimates of θ and σ^2 are obtained by differentiating the logarithm of likelihood (5) with respect to μ and σ^2 and setting the results equal to zero:

$$\text{(6)}\quad \frac{\partial \log L_{\underset{\sim}{b},\phi}}{\partial\mu} = \Sigma \frac{N_t}{f(t)} \; \frac{\partial f(t)}{\partial\mu} = 0$$

and

$$\text{(7)}\quad \frac{\partial \log L_{\underset{\sim}{b},\phi}}{\partial\sigma^2} = \Sigma \frac{N_t}{f(t)} \; \frac{\partial f(t)}{\partial\sigma^2} = 0 \; .$$

The derivatives equal

$$\text{(8)}\quad \frac{\partial f(t)}{\partial\mu} = \int \left[(\theta-\mu)/\sigma^2\right] \gamma_t \; \exp(t\theta) \left\{\prod_{i=1}^{n} \left[1+\exp(\theta-b_i)\right]\right\}^{-1} \phi(\theta)d\theta$$

and

$$\text{(9)}\quad \frac{\partial f(t)}{\partial\sigma^2} = \int \tfrac{1}{2}\left[(\theta-\mu)^2/\sigma^4 - 1/\sigma^2\right] \gamma_t \; \exp(t\theta) \left\{\prod_{i=1}^{n} \left[1+\exp(\theta-b_i)\right]\right\}^{-1} \phi(\theta)d\theta$$

From (4), (8) and (9) it is clear that the elementary symmetric functions which do not contain θ, do not have to be computed. Equations (6) and (7) can be solved iteratively for μ and σ^2 by the Newton-Raphson method; in order to be able to compute the derivatives the integrals have to be approximated by sums. Andersen and Madsen also presents a likelihood ratio test for fit.

Point and interval estimates for the person parameters

The likelihood equation for the estimation of θ, given a score equal to t, equals

$$\text{(10)}\quad h(\theta) = \frac{\partial \log L_{\underset{\sim}{b}}(t|\theta)}{\partial\theta} = t - \Sigma \exp(\theta-b_i) \left[1+\exp(\theta-b_i)\right]^{-1} = 0$$

A person parameter estimate $\hat{\theta}_t$ is obtained by solving (10) interatively for θ. An interval estimate of θ - under the assumption of approximate normality of the distribution of the likelihood estimate $\hat{\theta}$ and the assumption that the information function is fairly constant in the neighbourhood of the likelihood estimate - can be obtained from the information function through

$$(11) \quad \operatorname{var}(\hat{\theta}) \approx I^{-1}(\hat{\theta}) = \left\{ \sum_{i=1}^{n} \frac{\exp(\hat{\theta}-b_i)}{1+\exp(\hat{\theta}-b_i)} \right\}^{-1} .$$

An empirical Bayes point estimate of θ is the posterior mean

$$(12) \quad \bar{\theta}_t = E(\theta|t) = \int \theta f(\theta|t)d\theta$$

where $f(\theta|t)$ is obtained from (3) replacing $g(\theta)$ by the estimated population distribution. The posterior variance can be computed as

$$(13) \quad \operatorname{var}(\theta|t) = E(\theta^2|t)-E^2(\theta|t) = \int \theta^2 f(\theta|t)d\theta - \bar{\theta}^2_t .$$

Under the assumption that the posterior distribution can be approximated by a normal distribution (12) and (13) can be used in order to obtain a posterior confidence interval for θ.

Equations (12) and (13) also can be used for the estimation of μ and σ^2 by means of a procedure due to Sanathanan and Blumenthal (1978). From the equation one obtains

$$(14) \quad \hat{\mu} = \sum_{t=0}^{n} (N_{t}/_{N}) \, E(\theta|t)$$

and

$$(15) \quad \hat{\sigma}^2 + \hat{\mu}^2 = \Sigma(N_{t}/_{N}) \, E(\theta^2|t).$$

Choosing initial values for μ and σ^2, $E(\theta|t)$ and $E(\theta^2|t)$ can be computed for $t=0,\ldots,n$. Through Equations (14) and (15) one obtains new values for μ and σ^2. One can repeat the procedure until convergence is obtained. As a by-product one obtains the posterior means and variances. The algorithm is simple in that no first and second order derivatives of $f(t)$ have to be derived, but it also is computationally slow.

Instead of the posterior mean, the posterior mode $\tilde{\theta}_t$ could be used as a point estimate of θ. The mode can be obtained by solving

$$(16)\quad k(\theta) = r - \Sigma\left\{\exp(\theta-b_i)\left[1+\exp(\theta-b_i)\right]^{-1}\right\} - \sigma^{-2}(\theta-\mu) = 0$$

for θ; the first part of Equation (16) equals the likelihood equation, the second part stems from the normal population distribution. The modal estimate is obtained by iteratively solving (16) for θ. One may use the maximum likelihood estimate $\hat{\theta}$ as a starting value. Using a two-term Taylor expansion of (16) one then obtains

$$(17)\quad k(\theta) \simeq k(\hat{\theta})+(\theta-\hat{\theta})\left(\frac{\partial k}{\partial\theta}\right)_{\theta=\hat{\theta}}$$

$$= -\frac{(\hat{\theta}-\mu)}{\sigma^2} + (\hat{\theta}-\theta)\left[I(\hat{\theta}) + \frac{1}{\sigma^2}\right] = 0$$

due to the fact that $h(\hat{\theta})$ (cf. Equation (10)) equals zero. Equation (17) can be rewritten as

$$(18)\quad \tilde{\theta} = \rho_{\hat{\theta}}\,\hat{\theta} + (1-\rho_{\hat{\theta}})\,\mu$$

where

$$\rho_{\hat{\theta}} = \sigma^2\left[\sigma^2 + I^{-1}(\hat{\theta})\right]^{-1}.$$

Equation (18) clearly indicates that the empirical Bayes estimates are regressed to the mean; Equation (18) is a Kelley-formula for the estimation of θ. Estimate (18), being the estimated $\tilde{\theta}$ after the first iteration, may of course still differ to a certain extent from the final modal estimate $\tilde{\theta}$.
Also for the modal estimate $\tilde{\theta}$ an interval estimate can be obtained using

$$(19)\quad \hat{\text{var}}(\theta|t) = \left[I(\tilde{\theta}) + \sigma^{-2}\right]^{-1}.$$

A simulation

In order to demonstrate the feasibility and usefulness of empirical Bayes estimation within the context of the Rasch model, a simulation study was performed. It was assumed that the population distribution was N(0,1). Further, the hypothetical test was assumed to consist of

twenty items, with all item parameters equal to zero.

Three hundred person parameters were randomly generated from the distribution N(0,1). Next, for each hypothetical person p a score pattern was generated with a probability of a correct answer $\exp(\theta_p)\left[1+\exp(\theta_p)\right]^{-1}$ for each item. The 300 x 20 data matrix was analyzed using the unconditional procedure for the estimation of person and item parameters from BICAL, written by Wright and Mead (cf. Wright and Stone, 1979). The procedure contains a correction for bias in the item parameter estimates arising in the situation where both sets of parameters are estimated simultaneously; this bias was proved by Andersen (1973) for the case of two items.

The item parameter estimates, based on the 293 cases with scores $1 \leq x \leq 19$, ranged from -.113 to .182. Clearly regressed item parameter estimates have more optimal characteristics even if one does not invoke the concept of a distribution from which items are randomly sampled (Efron and Morris, 1975). In this paper we will, however, make use of the ordinary maximum likelihood estimates of the item parameters. The estimated population mean and variance, using the Andersen and Madsen procedure, are $\hat{\mu} = 0.047$ and $\hat{\sigma}^2 = 1.108$. The person parameter estimates $\hat{\theta}$, $\bar{\theta}$ and $\tilde{\theta}$ are given in Table 1. The posterior means were also computed by means of the Sanathanan and Blumenthal procedure; they are not presented here while they were virtually identical with the estimates from the Andersen and Madsen procedure.

For x equal to zero or twenty no maximum likelihood estimate is possible. In order to compare the maximum likelihood and the empirical Bayes estimate, 'maximum likelihood' estimates for scores t = 0 and t = 20 were computed by substitution of scores t = 0.5 and t = 19.5 respectively in the likelihood equation.

Tabel 1. Person parameter estimates

score	$\hat{\theta}$	$\bar{\theta}$	$\tilde{\theta}$
0	-3.67	-2.29	-2.21
1	-2.79	-1.93	-1.87
2	-2.08	-1.63	-1.58
3	-1.65	-1.37	-1.33
4	-1.32	-1.14	-1.10
5	-1.04	-0.92	-0.90
6	-0.80	-0.72	-0.70
7	-0.59	-0.53	-0.52
8	-0.39	-0.35	-0.34
9	-0.19	-0.17	-0.17
10	0	0.01	0.0
11	0.19	0.18	0.17
12	0.38	0.36	0.34
13	0.59	0.55	0.52
14	0.80	0.74	0.70
15	1.04	0.94	0.90
16	1.32	1.16	1.10
17	1.65	1.39	1.33
18	2.09	1.66	1.58
19	2.79	1.96	1.87
20	3.67	2.32	2.21

It is clear from Table 1 that the empirical Bayes point estimates are regressed to the mean, the posterior mode slightly more than the posterior mean estimates. Further, the regression is strongest for extreme observed scores. The confidence intervals based on the empirical Bayes approach, are smaller than the intervals based on the information function, as can be seen from Table 2. Again, the empirical Bayes approach has the largest effect for extreme scores.

Table 2. Estimated variances

score	$I^{-1}(\hat{\theta})$	var $(\theta\|x)$
0	2.05	0.38
1	0.91	0.32
2	0.51	0.28
3	0.37	0.25
4	0.30	0.22
5	0.26	0.21
6	0.23	0.19
7	0.22	0.19
8	0.21	0.18
9	0.20	0.18
10	0.20	0.18
11	0.20	0.18
12	0.21	0.18
13	0.22	0.19
14	0.23	0.20
15	0.26	0.21
16	0.30	0.23
17	0.37	0.25
18	0.51	0.28
19	0.91	0.33
20	2.05	0.39

In order to compare the effectiveness of the three different point estimates, the mean squared error loss in the sample was computed. This was done for all 300 cases and for the subgroup with scores 1 thru 19. The results are given in Table 3. Between brackets the results of two other simulations with 300 cases are presented. The empirical Bayes estimators

Table 3. Mean squared error loss

estimator	all cases (N=300)
$\hat{\theta}$	.303 (.250, .286)
$\bar{\theta}$	.225 (.180, .209)
$\tilde{\theta}$	.227 (.182, .213)
	cases with scores 1 thru 19
$\hat{\theta}$	.276 (.221, .244)
$\bar{\theta}$	.224 (.181, .196)
$\tilde{\theta}$	.224 (.184, .200)

outweigh the maximum likelihood estimator, with a slight advantage for the posterior mean as expected with the criterion of mean squared error loss. It is clear that the advantage of the empirical Bayes estimators is due in large part to the strong regression for extreme $\hat{\theta}$.

Discussion

Person parameter estimates which are regressed to the mean are on the average more accurate than ordinary maximum likelihood estimates. In this paper this is demonstrated for three samples of 300 cases within the context of the Rasch model. The empirical Bayes estimates varied more than the likelihood estimates, due to the fact that not only item parameters, but also population distribution parameters had to be estimated from the sample data, the largest difference being 0.0? for a score equal to 20. In applications of course new samples should be combined with old samples in order to improve all parameter estimates. The results were based on a computer simulation for which the model and distributional assumptions were known to be correct. For real data - apart from the question of fit of the Rasch model - there is a question of appropriateness of the distributional assumptions. It remains to be examined how robust the results are when the normality assumption is violated. The normality assumption can be tested by means of the test of fit proposed by Andersen and Madsen. If the normality assumption turns out to be inadequate, other distributions may be fitted to the data using a similar approach as proposed in this paper.

Interestingly, the fact that regressed estimates of person parameters are more efficient than the maximum likelihood estimator, is similar to regression effects in classical test theory. Regression effects will be larger with short tests and/or a small population variance

(an example of the latter case is given by Wood, 1978). It is clear that the empirical Bayes estimates of person parameters are important for many applications. In studies of change, e.g., where different groups of persons differ in initial estimated abilities, it is not possible to obtain sound conclusions if one does not account for regression effects.

References

Andersen, E.B. Conditional inference for multiple-choice questionnaires. British Journal of Mathematical and Statistical Psychology, 1973, 26, 31-44.

Andersen, E. and Madsen, M. Estimating the parameters of the latent population distribution. Psychometrika, 1977, 42, 357-374.

Birnbaum, A. Statistical theory for logistic mental test models with a prior distribution of ability. Journal of Mathematical Psychology, 1969, 6, 258-276.

Bock, R.D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 1972, 37, 29-51.

Bock, R.D. and Lieberman, M. Fitting a response model for n dichotomously scored items. Psychometrika, 1970, 35, 179-197.

Efron, B. and Morris, C. Data analysis using Stein's estimator and its generalizations. Journal of the American Statistical Association, 1975, 70, 311-319.

Jensema, C.J. An application of latent trait mental test theory. British Journal of Mathematical and Statistical Psychology, 1974, 27, 29-48.

Leonard, T. Bayesian methods for binomial data. Biometrika, 1972, 59, 581-589.

Meredith, W. and Kearns, J. Empirical Bayes point estimates of latent trait scores without knowledge of the trait distribution. Psychometrika, 1973, 38, 533-554

Owen, R.J. A Bayesian approach to tailored testing. Research Bulletin, 69-92. Princeton, N.J.: Educational Testing Service.

Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. Psychometrika, 1973, 38, 221-233.

Sanathanan, L. and Blumenthal, S. The logistic model and estimation of latent structure. Journal of the American Statistical Association, 1978, 73, 794-799.

Urry, V.W. Ancillary estimators for the item parameters of mental test models. In Computers and testing: steps toward the inevitable conquest. Research Section Personnel Research and Development Center. U.S. Civil Service Commission, Washington, D.C. September, 1976.

Wood, R. Fitting the Rasch model - a heady tale. British Journal of Mathematical and Statistical Psychology, 1978, 31, 27-32.

Wright, B.D. and Stone, M.H. Best test design. Chicago: Mesa Press, 1979.